The Problem

Given a population of size $N$, how long does it take until $\tau_k$ the first time we have an individual with a prespecified sequence of $k$ mutations?

- Initially all individuals are type 0.
- Each individual is subject to replacement at rate 1.
- A copy is made of an individual chosen at random from the population.
- Type $j-1$ mutates to type $j$ at rate $u_j$.

The Limits of Darwinism

The malaria parasite *Plasmodium falciparum* has evolved resistance to chloroquine. This is due to two amino acid altering substitutions in PfCRT. Michael Behe in his book *The Edge of Evolution* calls such an event a chloroquine complexity cluster, or CCC. He concludes:

“There are 5000 species of modern mammals. If each species had an average of a million members and if a new generation appeared every year, and if this went on for two hundred million years, the likelihood of a single CCC appearing in the whole bunch over that entire time would only be 1 in a hundred.”
**References for k=2 result**


---

**Theorem 1.** If $N u_1 \to 0$ and $N \sqrt{u_2} \to \infty$

$$P(\tau_2 > t / N u_1 \sqrt{u_2}) \to e^{-t}$$

10,000 simulations of $n = 10^3$, $u_1 = 10^{-4}$, $\sqrt{u_2} = 10^{-2}$

---

**Behe is wrong**

If $N = 10^6$, $u_1 = u_2 = 10^{-9}$, waiting time is exponential $10^{7.5} = 31.6$ million years for one prespecified pair of mutations in one species.

“There are 5000 species of modern mammals. If each species had an average of a million members and if a new generation appeared every year, and if this went on for two hundred million years, the likelihood of a single CCC appearing in the whole bunch over that entire time would only be 1 in a hundred.”

$u_2 \to \sqrt{u_2}$ is a factor of 31,600

---

**Stochastic tunneling**

---

**Idea of Proof**

Since 1’s mutate to 2’s at rate $u_2$, $\tau_2$ will occur when there have been $O(1/u_2)$ births of individuals of type 1.

The number of 1’s is roughly a symmetric random walk, so $\tau_2$ will occur when the number of 1’s reaches $O(1/\sqrt{u_2})$.

$N \gg 1/\sqrt{u_2}$ guarantees that up to $\tau_2$ the number of 1’s is $o(N)$, so 1 mutations occur at rate $N u_1$.

The waiting time from the 1 mutation until the 2 mutant appears is of order $1/\sqrt{u_2}$. For this to be much smaller than the overall waiting time $1/N u_1 \sqrt{u_2}$ we need $N u_1 \ll 1$.

---

**A few details**

Consider the multitype branching process in which individuals die at rate 1, give birth to a new individual of the same type at rate 1, and individuals mutate from type $j-1$ to type $j$ at rate $u_j$.

The probability $q$ that an individual of type 1 eventually has a descendant of type 2 satisfies

$$q = \frac{1}{2 + u_2} (2q - q^2) + \frac{u_2}{2 + u_2}$$

$$0 = q^2 + u_2 q - u_2$$

$$q = \frac{-u_2 + \sqrt{u_2^2 + 4u_2}}{2} \sim \sqrt{u_2}$$
The probability that an individual of type 1 eventually has a descendant of type 2, \( \sim \sqrt{u_2} \).

If there were always \( N \) individuals of type 0, 1 mutants occur at times of a Poisson process with rate \( Nu_1 \). The time \( \sigma_2 \) of the birth of the 1 individual that has a 2 descendant will be exponential with rate \( Nu_1 \sqrt{u_2} \).

If \( Nu_1 \ll 1 \), \( \tau_2 - \sigma_2 = O(1/\sqrt{u_2}) = o(1/Nu_1 \sqrt{u_2}) \).

**If we wait for fixation**, replace \( u_2 \) by \( u_2 \beta \), where \( \beta \) = fixation probability. Small surprise is time is only increased by \( 1/\sqrt{\beta} \).

**If 1’s are mildly deleterious**, which means fitness \( 1 - \rho/\sqrt{u_2} \), instead of the usual \( 1 - O(1/N) \), time is increased by \( 1/R \) where \( R = (-\rho + \sqrt{\Delta + \rho^2})/2 \)

When \( Nu_1 \neq 0 \)

**Theorem 2.** Suppose that \( Nu_1 \to \lambda \in (0, \infty) \), \( u_2 \to 0 \), and \( N \sqrt{u_2} \to \infty \) as \( N \to \infty \). Then \( \tau_2 \cdot Nu_1 \sqrt{u_2} \) converges to a limit that has density function

\[
 f_2(t) = h(t) \exp \left( - \int_0^t h(s) \, ds \right) \quad \text{where} \quad h(s) = \frac{1 - e^{-2s/\lambda}}{1 + e^{-2s/\lambda}}
\]

**Sketch of proof**

Let \( Q_1 \) be the law of the process starting from a single 1 and modified to have no further 1 mutations. Let \( g_2(t) = Q_1(\tau_2 \leq t) \).

\[
g_2(t) = -u_2 g_2(t) - g_2(t)^2 + u_2
\]

Solve the ODE and then compute

\[
P(\tau_2 \leq t) = 1 - \exp \left( -Nu_1 \int_0^t g(s) \, ds \right)
\]


**Drosophila**

Suppose a transcription factor binding site consists of 10 nucleotides. Taking \( 10^{-8} \) as mutation rate, \( u_1 = 10^{-7} \) and \( u_2 = (1/3) \times 10^{-8} \).

\( N = 5 \times 10^6 \) chromosomes, so waiting time has mean \( 1/Nu_1 \sqrt{u_2} = 34,600 \) generations or 3,460 years assuming 10 generations per year.

\( Nu_1 \) is not small, but Theorem 2 and simulations suggest this adds 25% to total = 4,325 years.

In neutral case \( \beta = 1/2N \) increasing time by a factor of \( 1/\sqrt{\beta} = 2200 \) to 9 million years.

If two fitnesses are \( 1 - 10^{-4} \) and \( 1 + 10^{-4} \) answer is roughly 1 million years.

10,000 simulations of \( N = 10^3 \), \( u_1 = 10^{-3} \), \( \sqrt{u_2} = 10^{-2} \). The exponential with mean \( 1/Nu_1 \sqrt{u_2} \) is given by the thin line. The approximation from Theorem 2 by the thick line.
Probability type $j$ has a type $k$ descendant.
\[ \sim r_{j,k} = u_{j+1}^{1/2} u_{j+2}^{1/4} \cdots u_k^{1/2^{k-j}} \quad \text{for} \ 1 \leq j < k \]

**Theorem 3.** Let $k \geq 2$. Suppose that:

(i) $N_u \to 0$.
(ii) For $j = 1, \ldots, k-1$, $u_{j+1}/u_j > b_j$ for all $N$.
(iii) There is an $a > 0$ so that $N^2 u_k \to 0$.
(iv) $N_{r_{1,k}} \to \infty$.

Then for all $t > 0$, $\lim_{N \to \infty} P(\tau_k > t/N_u r_{1,k}) = \exp(-t)$.

### Ideas in Proof

In the branching process, the probability a type $j$ has a type $k$ descendant.

\[ p_{j,k} = \frac{1}{2 + u_{j+1}}(2p_{j,k} - p_{j,k}^2) + \frac{u_{j+1}}{2 + u_{j+1}}p_{j+1,k}. \]

Solving gives

\[ p_{j,k} = \frac{-u_{j+1} + \sqrt{u_{j+1}^2 + 4u_{j+1}p_{j+1,k}}}{2}. \]

Using (ii) and (iii) we conclude $p_{j,k} \sim r_{j,k}$

### When $N_{r_{1,k}} \not\to \infty$

Fixation of 1 before $\tau_k$ and stochastic tunneling each have positive probability.

**Theorem 4.** Let $k \geq 2$. Assume

(i) $N_u \to 0$.
(ii) For $j = 1, \ldots, k-1$, $u_{j+1}/u_j > b_j$ for all $N$.
(iii) There is an $a > 0$ so that $N^2 u_k \to 0$.
(iv) $(N_{r_{1,k}})^2 \to \gamma > 0$, and we let

\[ \alpha = \sum_{k=1}^{\infty} \frac{\gamma^k}{(k-1)!(k-1)!} / \sum_{k=1}^{\infty} \frac{\gamma^k}{k!(k-1)!} > 1 \]

then for all $t > 0$, $\lim_{N \to \infty} P(u_1 \tau_k > t) = \exp(-\alpha t)$.

10,000 simulations with $N = 10^3$, $u_1 = 10^{-4}$, $\sqrt{v} = 10^{-3}$
\[ \gamma = 1, \alpha = 1.433 \]

**Explanation of the Conditions**

\[ r_{j,k} = u_{j+1}^{1/2} u_{j+2}^{1/4} \cdots u_k^{1/2^{k-j}} \quad \text{for} \ 1 \leq j < k \]

(i) $N_u \to 0$ implies we can ignore $\tau_k - \sigma_k$, where $\sigma_k$ is the birth time of the type 1 with a type $k$ descedant.
(ii) $N_{r_{1,k}} \to \infty$ guarantees the number of mutants stays $o(N)$.
(iii) $N^2 u_k \to 0$ for some $a > 0$. Mutation rates can’t be too big.

If $k = 4$ and $u_j = \mu$ then there are $\mu^{-1/2} 3$’s; $\mu^{-3/4} 2$’s; $\mu^{-7/8} 1$’s.

Processes live on different time scales.

Use induction to reduce to two type case.
Let $X_j(t)$ be the number of type $j$ at time $t$.

If $X_1(0) = N\epsilon$ then $N^{-1}X_1(Nt) \rightarrow Z_t$ where $Z_t$ is the Wright-Fisher diffusion process with infinitesimal generator $x(1-x)d^2/dx^2$.

When $X_1(Nt) = N\epsilon$, mutations to type 2 that eventually lead to a type $k$ individual occur at rate approximately

$$N \cdot N\epsilon \cdot u_{2, k} \sim N^2 \epsilon^2 \rightarrow \gamma \epsilon,$$

so if we let $u(x)$ be the probability that the process $Z_t$ hits 0 before reaching 1 or generating a type $m$ mutation, then $u(x)$ satisfies

$$x(1-x)u''(x) - \gamma xu(x) = 0, \quad u(0) = 1, \quad u(1) = 0$$

The constant $\alpha = \lim_{\epsilon \rightarrow 0} (1 - u(\epsilon))/\epsilon$.

### Small time behavior

Our results are appropriate for the regulatory sequence application since one is interested in the typical amount of time that the process takes.

However, most cancers occur in less than 1% of the population so we are looking at the lower tail of the distribution. Let $g_j(t) = Q_1(\tau_j \leq t)$ where $Q_1$ is the probability for the branching process started with one type 1. In the case $u_j \equiv \mu$

$$g_j(t) = \mu g_j(t) - (1 - \mu)g_j(t)^2 - 2\mu g_j(t)$$

One can inductively solve the differential equations and finds

If $t \ll \mu^{-1/2}$ then $g_j(t) \approx \mu^{k-1}k^{-1}/(k - 1)!$